Systematic tracking of altered modules identifies disrupted pathways in teratozoospermia

Z.Q. Huang[1,2], G.X. Wang[3], X.L. Jiang[4], E.P. Tian[2], W.L. Yao[2] and T. Zeng[5]

[1]School of Medicine, Nanchang University, Nanchang, China
[2]Department of Urinary Surgery, Nanchang Reproductive Hospital, Nanchang, China
[3]Department of Urinary Surgery, First Affiliated Hospital of Nanchang University, Nanchang, China
[4]Department of Reproductive Medicine, Nanchang Reproductive Hospital, Nanchang, China
[5]Department of Urinary Surgery, Jiangxi Province Peoples Hospital, Nanchang, China

Corresponding author: Z.Q. Huang
E-mail: huangzhiqian1234@163.com

Genet. Mol. Res. 15 (2): gmr.15027514
Received August 25, 2015
Accepted December 3, 2015
Published April 25, 2016
DOI http://dx.doi.org/10.4238/gmr.15027514

ABSTRACT. The objective of this study was to identify disrupted pathways in teratozoospermia by systematically tracking dysregulated modules in reweighted protein-protein interaction (PPI) networks. We inferred and reweighted the PPI networks of normal and teratozoospermia groups based on Spearman correlation coefficients. Modules in the PPI networks were explored via a clique-merging algorithm and altered modules were identified based on maximum weight bipartite matching. Furthermore, pathway-enrichment analyses of genes in altered modules were performed by Database for Annotation, Visualization, and Integrated Discovery (DAVID) to illuminate the biological pathways in teratozoospermia. A total of 20,102 genes were screened from the expression profile. We explored

2406 and 2101 modules in normal and disease PPI networks, respectively. Moreover, we obtained 875 altered modules by comparing modules in normal and teratozoospermia PPI networks. At $P < 0.01$, the genes involved in 2855 interactions with score changes >1 were mainly enriched in 66 pathways and the genes in altered modules were enriched in 71 pathways. The activity genes (missed and added genes in the disrupted modules) were enriched in 41 common pathways. There were 36 mutual enriched pathways under the five different conditions. Moreover, the cell cycle pathway was disrupted in the first 10 pathways of each condition. This study provides a powerful biomarker discovery platform to better understand the progression of teratozoospermia by systematically tracking dysregulated modules. This method uncovered potential diagnostic and therapeutic targets of teratozoospermia. This information might lead to improved monitoring and treatment of teratozoospermia.

Key words: Teratozoospermia; Protein-protein interaction network; Altered modules; Disrupted pathways

INTRODUCTION

Teratozoospermia is a major factor contributing to male infertility, but its molecular pathogenesis is usually unknown. It is characterized by the presence of spermatozoa with abnormal morphology in over 85% in sperm (De Braekeleer et al., 2015). These abnormalities may include heads that are large, small, tapered, or pyriform or tails that are abnormally shaped. The causes of poor morphology can be related to lifestyle and habits, such as smoking and toxin exposure. Teratozoospermia, including the globozoospermia type (Egashira et al., 2009), may be treated by intracytoplasmic sperm injection that injects sperm directly into the egg (French et al., 2010) for optimal success rates.

To gain a better understanding of the genetic basis of teratozoospermia, previous studies confirmed several genes correlated with the condition. Lin et al. (2012) provided the first causal link between a *SEPTIN12* genetic variant and male infertility with a distinctive sperm pathology. Meanwhile, Kusz-Zamelczyk et al. (2013) were the first to describe the association between *NANOS1* gene mutations with human severe oligoasthenoteratozoospermia. Recent reports also suggest that mutations or deletions in *AURKC*, *SPATA16*, *PICK1*, and *DPY19L2* are responsible for teratozoospermia (De Braekeleer et al., 2015). However, the genomic or gene expression variation contributing to teratozoospermia is still unknown.

Recent developments in high-throughput experimental technologies have allowed the emergence of large amounts of protein-protein interaction (PPI) data that make it possible to study proteins on a systematic level (Jordán et al., 2012). We can focus on genes or modules associated with disease to understand the mechanism of disease through a PPI network. However, it was originally difficult for PPI data to predict complexes because of its high false-positive and false-negative rates (Wu et al., 2012). Now many computational approaches have been proposed that are capable of assessing the reliability of protein interaction data. Liu et al. (2008) identified an iterative scoring method that could evaluate reliability and predict new interactions better than other methods. It is crucial to study the behavior of modules across specific conditions in a controlled

manner to understand the modus operandi of disease mechanisms and to implicate novel genes (Srihari and Ragan, 2013), since some important genes may not be identifiable through their own behavior when considered in conjunction with other genes (i.e., as modules). Many human genes have not yet been assigned to definitive pathways, so a scoring pathway based on module analysis becomes a more reliable approach.

To further reveal the mechanism of teratozoospermia, we systematically tracked the altered modules of reweighted PPI networks to identify disturbed pathways between normal and teratozoospermia patients. We inferred normal and teratozoospermia PPI networks based on Spearman correlation coefficients (SCCs) and explored modules from the PPI networks using a clique-merging algorithm. Moreover, we compared these modules to identify altered modules. Finally, we performed pathway-enrichment analysis for genes in altered modules based on Database for Annotation, Visualization, and Integrated Discovery (DAVID). This study provides a broad application platform for the identification of disrupted biological pathways associated with teratozoospermia and for the discovery of potential new molecular targets for the development of more effective therapies to treat teratozoospermia.

MATERIAL AND METHODS

Data recruitment and preprocessing

Microarray expression data

Several microarray databases are now recognized, such as ArrayExpress and gene expression omnibus (GEO). These databases store abundant microarray experimental data that have been uploaded from various laboratories worldwide and made freely available for scientific research. Here we analyzed a teratozoospermia microarray expression profile, E-GEOD-6872, from the ArrayExpress database. E-GEOD-6872 contains 13 normal semen samples and 8 semen samples from teratozoospermic individuals. These expression data were derived from the Affymetrix GeneChip Human Genome U133A 2.0 platform.

The probe data were obtained via the expresso function from Bioconductor affy package to eliminate the influence of nonspecific hybridization. Background correction was performed by robust multichip average (RMA) (Irizarry et al., 2003), quantile-based algorithm was used to standardize the data (Bolstad et al., 2003), PM (perfect match)/MM (mismatch match) correction was applied by the microarray suite (MAS) method (Gautier et al., 2004), and Medianpolish (Irizarry et al., 2003) was conducted to summarize the probe data. The original probe set yielded 20,102 genes after filtering by featureFilter functions.

PPI data

We downloaded human protein interaction data comprised of 1,048,576 interactions from the database Search Tool for the Retrieval of Interacting Genes/Proteins (STRING, http://string-db.org/) that provided a comprehensive, yet quality-controlled collection of protein-protein associations for a large number of organisms (von Mering et al., 2005). The protein and gene name were transformed. We then removed self-loops and proteins without expression values. The constructed PPI network consisted of 8590 nodes and 53,975 highly correlated interactions with a

combined-score ≥0.8. The network containing of 8015 genes and 49,451 edges was established by taking the intersection of the 20,102 genes included in E-GEOD-6872 and the nodes in PPI network.

Construction of reweighted PPI network

We reweighted the gene interactions in the teratozoospermia and normal networks, to include SCC values between -1 and +1, which is a popular method to describe the interaction or edge strength of gene pairs. The SCC of every interaction was calculated by including all edges in the network. Positive SCCs indicate positive linear correlations between gene pairs in the normal and disease groups. Negative SCCs indicate negative correlations between gene pairs in the groups. Furthermore, we defined SCCs as weighted values of the interaction in the teratozoospermia network and normal network. Interactions with an absolute difference in SCCs >1 were selected for further analysis.

Identification of modules

We used the clique-merging method (Liu et al., 2009; Srihari and Leong, 2013) to identify the module from PPI networks. First, we sought all maximal cliques from the normal and disease networks based on the fast depth-first method. We then selected the maximal cliques with greater than five nodes and calculated their weighted interaction density. Lastly, we arranged the cliques in descending order. The score of a clique *C* was defined as its weighted interaction density:

$$\text{score}(C)=\frac{\sum_{u\in C,v\in c} w(u,v)}{|C|.(|C|-1)} \quad \text{(Equation 1)}$$

where $w(u, v)$ represents the weight of the interactions between *u* and *v*. If the density of two cliques was the same, the weighted density of the larger clique was likely to be higher than that of the smaller clique.

Many maximal cliques from a PPI network may overlap with others. Thus, highly overlapped cliques should be removed to reduce result size or merged to form bigger modules. The interconnectivity between two cliques was used to decide whether two overlapped cliques should be merged together or not. The interconnectivity between the non-overlapping part of C_i and C_j was calculated as follows:

$$\text{inter-score}(C_i,C_j)=\sqrt{\frac{\sum_{u\in(C_i-C_j)}\sum_{v\in C_j}(u,v)}{|C_i-C_j|.|C_j|}\cdot\frac{\sum_{u\in(C_j-C_i)}\sum_{v\in C_i}w(u,v)}{|C_j-C_i|.|C_i|}} \quad \text{(Equation 2)}$$

Given a set of cliques ranked in order of descending score, denoted as $\{C_1, C_2, ..., C_k\}$, the algorithm removed and merged highly overlapped cliques in the following steps. For each maximal clique C_i, if there was another maximal clique C_j, under the condition of $|C_i \cap C_j| / |C_j| > t_0$ (a predefined overlap threshold) (Srihari and Ragan, 2013), we checked the weighted interconnecting scores of both C_i and C_j. If the inter-score was greater than t_m (a predefined merge threshold) (Srihari and Ragan, 2013), C_i and C_j were combined into the module. Here, $t_0 = 0.5$ and $t_m = 0.25$.

Identification of altered modules

The node number statistics of each module were analyzed in the normal and disease groups. The module correlation density was calculated according to the SCC of edges in PPI network. $S = \{S_1, S_2 \ldots, S_n\}$ and $T = \{T_1, T_2 \ldots, T_m\}$ were the sets of modules identified from the normal and disease networks, respectively. For each $S_i \in S$, the module correlation density was measured as follows:

$$d_{cc}(S_i) = \frac{\sum_{p,q \in Si} PCC((p,q),N)}{\left(\frac{|S_i|}{2}\right)} \quad \text{(Equation 3)}$$

The correlation densities for disease modules T were calculated similarly.

A similarity graph M = (Vm, Em) was established based on the node for the module and the edge for two modules, where $V_M = \{ S \cup T \}$, $E_M = \cup\{(S_i, T_j): J(S_i, T_j) \geq t_J$, and $\Delta_{CC}(S_i, T_j) \geq \delta\}$, where $J(S_i, T_j) = |S_i \cap T_j|/|S_i \cup T_j|$ was the Jaccard similarity (J) and $\Delta_{CC}(S_i, T_j) = |d_c(S_i)\text{-}d_c(T_j)|$ was the differential correlation density (Δ_{CC}) between S_i and T_j (Srihari and Ragan, 2013). After identifying altered modules by maximum weight bipartite matching (Gabow, 1976) in M, we inferred genes involved in teratozoospermia as $\Gamma = \{g: g \in S_i \cup T_j, (S_i, T_j) \in \Upsilon(S, T)\}$ ranked in descending order of $\Delta_{CC}(S_i, T_j)$. To identify altered modules, we matched normal and teratozoospermia modules by setting a high t_J, ensuring that the modules either had the same gene composition or had lost or gained only a few genes. J = 1 represented disrupted modules with the same gene composition. The altered modules whose $t_J \geq 2/3$ and $\Delta_{CC} \geq 0.05$ were defined as distinct modules.

Comparison of gene composition in altered modules

Differences between modules are ultimately due to changes in edge or gene composition. Hence, we performed the gene composition analysis in altered modules. Compared with each module in normal condition, we selected the altered genes in the corresponding disrupted module, including missing genes (genes appeared in normal modules but not in disease ones) and added genes (genes appeared in disease modules but not in normal ones). Interestingly, some genes maybe missing (or added) in one condition, but also added (or missing) in another condition. These both added and missing genes in disrupted modules were also analyzed.

Pathway-enrichment analysis of genes in altered modules

We conducted Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway-enrichment analysis for genes in altered modules from normal and teratozoospermia patients to link genomic information with higher order functional information regarding cellular processes and standardized gene annotations (Kanehisa and Goto, 2000), based on DAVID (Huang et al., 2009). Pathway analyses of the active genes, including added genes, missed genes, and intersection genes were also conducted. Values of $P < 0.01$ indicated significant differences in this study.

RESULTS

Disruptions in the teratozoospermia PPI network

After data preprocessing, a total of 20,102 genes were screened from the expression profile. The PPI network displayed equal numbers of interactions (49,151) with average scores of 0.323 and 0.361 in the normal and disease condition, respectively. SCC distributions in the two networks are shown in Figure 1. The number of interactions in the teratozoospermia network was greater than in the normal network with score distributions of -1.0-0.5 and 0.6-1.0. Examining all interactions more carefully, we found that the scores of 29,156 interactions in the teratozoospermia network were lower than those in the normal network. On the contrary, the scores of 19,995 interactions were higher than scores in the normal network. We extracted interactions with score changes >1 (2855 interactions) for further analysis.

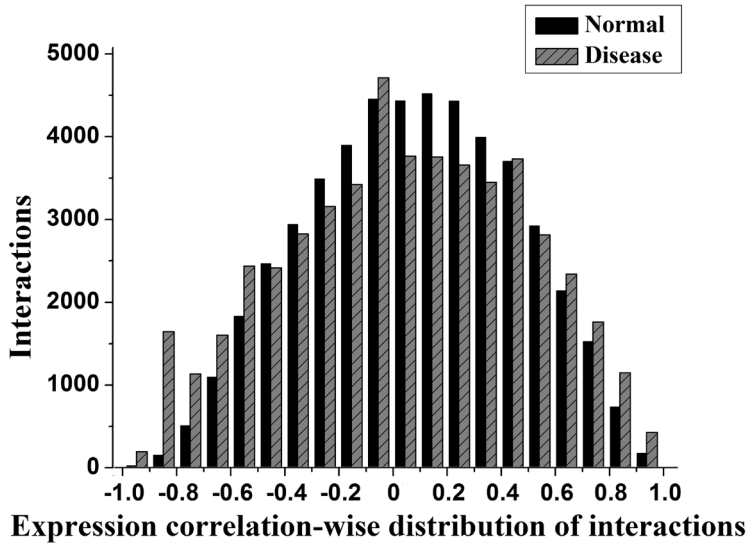

Figure 1. Expression correlation-wise distribution of interactions in the normal and disease groups.

KEGG pathway-enrichment analysis of genes involved in these 2855 interactions was performed. At $P < 0.01$, these genes were mainly enriched in 66 biological processes. Pathways in cancer ($P = 8.86^{-23}$) and cell cycle ($P = 2.32^{-18}$) were the most significant pathways.

Analyzing disruptions in teratozoospermia modules

A clique-merging algorithm was performed to identify altered modules from the normal and teratozoospermia PPI network. A total of 8405 maximal cliques were identified when considering cliques with greater than five nodes. We found 2406 and 2101 altered modules in the normal and disease PPI networks, respectively (Table 1). Meanwhile, the average module sizes were very similar across the two conditions. Disease modules showed an overall decrease in correlation

compared with the normal modules. Figure 2 shows the relationship between the numbers of modules and weighted density of modules. There were significant differences between the distribution of modules in normal and disease networks when the weighted density ranged from 0 to 0.3. Moreover, the distribution of modules in the disease network was higher under expression correlation values of -0.1-0.1 and lower under expression correlation values of 0.1-0.4 relative to that in normal condition.

Table 1. Properties of normal and disease modules.

Module set	No. of modules	Average module size	Correlation		
			Max	Avg	Min
Normal	2406	20.93 ± 18.51	0.46	0.08	-0.15
Disease	2101	19.43 ± 13.08	0.41	0.03	-0.19

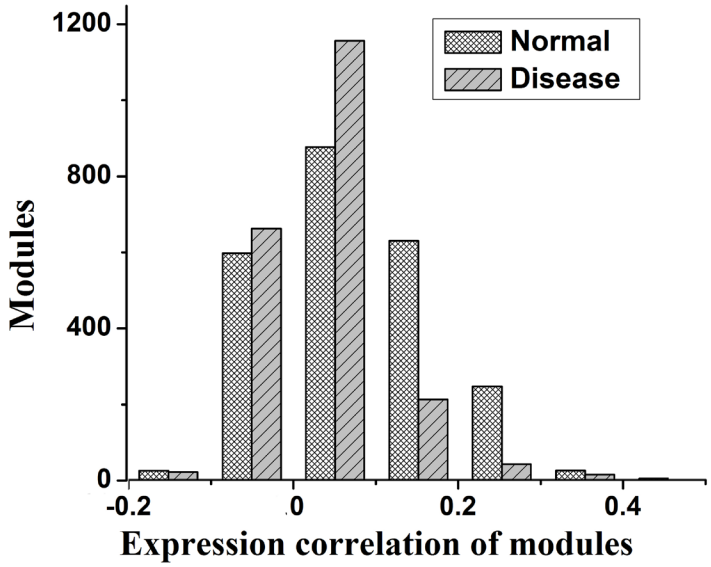

Figure 2. Correlation-wise distribution of modules in the normal and disease groups.

We obtained 875 altered modules at the threshold of t_j = 2/3 and Δ_{cc} = 0.05. Comparing with the normal condition, a total of 204 altered modules showed higher module correlation density in disease condition. There were 1615 genes in these altered modules. Pathway analysis was conducted for these genes. These genes were mainly enriched in 71 terms ($P < 0.01$), of which cell cycle ($P = 5.75^{-22}$) and pathways in cancer ($P = 2.97^{-19}$) were the top two disrupted pathways.

In-depth analyses of altered modules

There were 505 missed genes and 611 added genes, and 188 intersection genes. Pathway analyses based on these genes were conducted separately. At a threshold of $P < 0.01$, the added genes were enriched in 57 terms, the missed genes were enriched in 54 terms, and the intersection of the missed and added genes were enriched in 45 terms. A Venn diagram of three

sets of pathway terms shows that there were 41 common terms in the three conditions (Figure 3). The results of the five pathway analyses, a total of 36 mutual pathways, are shown in Table 2. Notably, the cell cycle pathway was disrupted in the first 10 pathways of each condition.

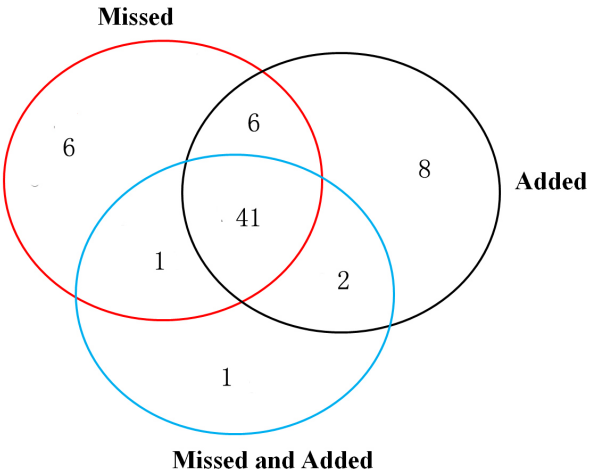

Figure 3. Venn diagram of pathway terms for missed genes, added genes, and the intersection of missed and added genes.

Table 2. Results of the mutual pathways under the five different conditions.

KEGG ID	Pathways
hsa05220	Chronic myeloid leukemia
hsa04664	Fc epsilon RI signaling pathway
hsa00590	Arachidonic acid metabolism
hsa00591	Linoleic acid metabolism
hsa04912	GnRH signaling pathway
hsa05215	Prostate cancer
hsa04010	MAPK signaling pathway
hsa04110	Cell cycle
hsa04270	Vascular smooth muscle contraction
hsa04062	Chemokine signaling pathway
hsa05211	Renal cell carcinoma
hsa05213	Endometrial cancer
hsa05223	Non-small cell lung cancer
hsa05210	Colorectal cancer
hsa05200	Pathways in cancer
hsa04510	Focal adhesion
hsa05221	Acute myeloid leukemia
hsa04370	VEGF signaling pathway
hsa04662	B cell receptor signaling pathway
hsa00980	Metabolism of xenobiotics by cytochrome P450
hsa05214	Glioma
hsa04730	Long-term depression
hsa04722	Neurotrophin signaling pathway
hsa04540	Gap junction
hsa04660	T cell receptor signaling pathway
hsa04810	Regulation of actin cytoskeleton
hsa00982	Drug metabolism
hsa04012	ErbB signaling pathway
hsa05212	Pancreatic cancer
hsa05219	Bladder cancer
hsa04914	Progesterone-mediated oocyte maturation
hsa04350	TGF-beta signaling pathway
hsa00230	Purine metabolism
hsa04080	Neuroactive ligand-receptor interaction
hsa03020	RNA polymerase
hsa03050	Proteasome

DISCUSSION

In order to identify disrupted pathways in teratozoospermia, this study performed a systematic analysis of altered modules from reweighted normal and teratozoospermia PPI networks based on SCCs. A total of 875 altered module pairs, including 1615 genes, were identified. We selected 611 added genes, 505 missed genes, and 188 intersection genes in disease modules compared to normal modules. We performed pathway analysis for the genes in altered modules, the active genes (the added and missed genes), and genes with an SCC change >1. Pathway-enrichment analysis revealed that the cell cycle pathway was commonly enriched in the first 10 pathways of each condition.

Recent rapid advances in high-throughput technologies have brought unprecedented opportunities for large-scale analyses of disease related molecular mechanisms. These technologies have transformed data analysis to reveal meaningful biological phenomena. Analytical approaches such as pathway analyses have become important and powerful tools to elucidate biological mechanisms underlying complex diseases. A majority of recent genome-related studies have been aimed at the verification of the enrichment pathways via the differentially expressed genes (Peukert et al., 1997). However, little research concerning disrupted pathways associated with disease based on human altered module analysis has been performed. The disrupted pathways analysis can better describe phenotypic differences from the viewpoint of modules in contrast to traditional DEG methods. Thus, pathway analysis based on modules offers a novel method to accurately predict mechanisms underlying diseases.

The cell cycle, or cell-division cycle, includes the series of events that take place in a cell leading to its division and duplication (Hirt, 2013). Cell cycle progression is ordered into dependent pathways in which the initiation of later events depends on the completion of earlier events (Hartwell and Weinert, 1989). Some studies have confirmed that many genes associated with cell cycle regulation, such as *AURKC* (El Kerch et al., 2011) and *ATM* (Orozco-Lucero et al., 2014), affected the occurrence of teratozoospermia and may be involved in the disease. These insights suggest molecular mechanisms for cellular transformation that may help identify potential targets for improving teratozoospermia therapies. Therefore, the results of this study were consistent with previous studies that found cell cycle to be the most significant pathway involved in teratozoospermia.

In this study, we successfully obtained some disrupted pathways associated with teratozoospermia that might yield biomarkers for diagnosis and treatment by systematically tracking the altered modules from reweighted normal and teratozoospermia PPI networks. Moreover, we concluded that cell cycle was the most significant pathway that may be closely related to the occurrence of teratozoospermia.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

We would like to thank Ji'nan Evidence Based Medicine Science-Technology Center for the translation assistance to the manuscript.

REFERENCES

Bolstad BM, Irizarry RA, Astrand M and Speed TP (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 19: 185-193.http://dx.doi.org/10.1093/bioinformatics/19.2.185

De Braekeleer M, Nguyen MH, Morel F and Perrin A (2015). Genetic aspects of monomorphic teratozoospermia: a review. *J. Assist. Reprod. Genet.* 32: 615-623.http://dx.doi.org/10.1007/s10815-015-0433-2

Egashira A, Murakami M, Haigo K, Horiuchi T, et al. (2009). A successful pregnancy and live birth after intracytoplasmic sperm injection with globozoospermic sperm and electrical oocyte activation. *Fertil. Steril.* 92: 2037.e5-2037.e9.http://dx.doi.org/10.1016/j.fertnstert.2009.08.013

El Kerch F, Lamzouri A, Laarabi FZ, Zahi M, et al. (2011). [Confirmation of the high prevalence in Morocco of the homozygous mutation c.144delC in the aurora kinase C gene (AURKC) in the teratozoospermia with large-headed spermatozoa]. *J. Gynecol. Obstet. Biol. Reprod. (Paris)* 40: 329-333.http://dx.doi.org/10.1016/j.jgyn.2010.09.003

French DB, Sabanegh ES, Jr., Goldfarb J and Desai N (2010). Does severe teratozoospermia affect blastocyst formation, live birth rate, and other clinical outcome parameters in ICSI cycles? *Fertil. Steril.* 93: 1097-1103.http://dx.doi.org/10.1016/j.fertnstert.2008.10.051

Gabow HN (1976). An efficient implementation of Edmonds' algorithm for maximum matching on graphs. *J. Assoc. Comput. Mach.* 23: 221-234. http://dx.doi.org/10.1145/321941.321942

Gautier L, Cope L, Bolstad BM and Irizarry RA (2004). affy - analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* 20: 307-315.http://dx.doi.org/10.1093/bioinformatics/btg405

Hartwell LH and Weinert TA (1989). Checkpoints: controls that ensure the order of cell cycle events. *Science* 246: 629-634. http://dx.doi.org/10.1126/science.2683079

Hirt BV (2013). Mathematical modelling of cell cycle and telomere dynamics. Ph.D. thesis, University of Nottingham.

Huang W, Sherman BT and Lempicki RA (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4: 44-57.http://dx.doi.org/10.1038/nprot.2008.211

Irizarry RA, Bolstad BM, Collin F, Cope LM, et al. (2003). Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res.* 31: e15.http://dx.doi.org/10.1093/nar/gng015

Jordán F, Nguyen TP and Liu WC (2012). Studying protein-protein interaction networks: a systems view on diseases. *Brief. Funct. Genomics* 11: 497-504.http://dx.doi.org/10.1093/bfgp/els035

Kanehisa M and Goto S (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28: 27-30.http://dx.doi.org/10.1093/nar/28.1.27

Kusz-Zamelczyk K, Sajek M, Spik A, Glazar R, et al. (2013). Mutations of NANOS1, a human homologue of the *Drosophila* morphogen, are associated with a lack of germ cells in testes or severe oligo-astheno-teratozoospermia. *J. Med. Genet.* 50: 187-193.http://dx.doi.org/10.1136/jmedgenet-2012-101230

Lin YH, Wang YY, Chen HI, Kuo YC, et al. (2012). SEPTIN12 genetic variants confer susceptibility to teratozoospermia. *PLoS One* 7: e34011.http://dx.doi.org/10.1371/journal.pone.0034011

Liu G, Li J and Wong L (2008). Assessing and predicting protein interactions using both local and global network topological metrics. *Genome Inform* 21: 138-149. PubMed

Liu G, Wong L and Chua HN (2009). Complex discovery from weighted PPI networks. *Bioinformatics* 25: 1891-1897.http://dx.doi.org/10.1093/bioinformatics/btp311

Orozco-Lucero E, Dufort I, Robert C and Sirard MA (2014). Rapidly cleaving bovine two-cell embryos have better developmental potential and a distinctive mRNA pattern. *Mol. Reprod. Dev.* 81: 31-41.http://dx.doi.org/10.1002/mrd.22278

Peukert K, Staller P, Schneider A, Carmichael G, et al. (1997). An alternative pathway for gene regulation by Myc. *EMBO J.* 16: 5672-5686.http://dx.doi.org/10.1093/emboj/16.18.5672

Srihari S and Leong HW (2013). A survey of computational methods for protein complex prediction from protein interaction networks. *J. Bioinf. Comput. Biol*. 11.

Srihari S and Ragan MA (2013). Systematic tracking of dysregulated modules identifies novel genes in cancer. *Bioinformatics* 29: 1553-1561.http://dx.doi.org/10.1093/bioinformatics/btt191

von Mering C, Jensen LJ, Snel B, Hooper SD, et al. (2005). STRING: known and predicted protein-protein associations, integrated and transferred across organisms. *Nucleic Acids Res.* 33: D433-D437.http://dx.doi.org/10.1093/nar/gki005

Wu C, Zhu J and Zhang X (2012). Integrating gene expression and protein-protein interaction network to prioritize cancer-associated genes. *BMC Bioinformatics* 13: 182.http://dx.doi.org/10.1186/1471-2105-13-182